FUNDAMENTAL IDEAS FROM RORU83

Jesus Cerquides
IIIA-CSIC

10/10/2020 ALGO group

WHY DO WE NEED TO HEAR THIS TALK?

Move to the notebook…

¿RORU83?

Biometrika (1983), **70**, 1, *pp.* 41–55
Printed in Great Britain

The central role of the propensity score in observational studies for causal effects

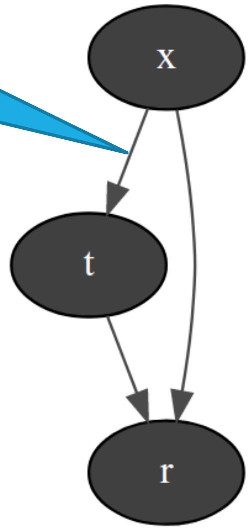
By PAUL R. ROSENBAUM
Departments of Statistics and Human Oncology, University of Wisconsin, Madison, Wisconsin, U.S.A.

AND DONALD B. RUBIN
University of Chicago, Chicago, Illinois, U.S.A.

The central role of the propensity score in observational studies for causal effects
PR Rosenbaum, DB Rubin - Biometrika, 1983 - academic.oup.com
The propensity score is the conditional probability of assignment to a particular treatment given a vector of observed covariates. Both large and small sample theory show that adjustment for the scalar propensity score is sufficient to remove bias due to all observed covariates. Applications include:(i) matched sampling on the univariate propensity score, which is a generalization of discriminant matching,(ii) multivariate adjustment by subclassification on the propensity score where the same subclasses are used to estimate ...

APPROACH TO CAUSAL INFERENCE

The problem is brought in by this dependence



N units

2 treatments {0,1}

Stable unit-treatment value assumption

x_i: Observed pretreatment measurements

z_i: Treatment assigned to unit i

$r_{t,i}$: Response of unit i to treatment t

x_1	$z_1 = 1$	$r_{0,1}$	$r_{1,1}$
x_2	$z_2 = 0$	$r_{0,2}$	$r_{1,2}$
x_3	$z_3 = 0$	$r_{0,3}$	$r_{1,3}$
x_4	$z_4 = 1$	$r_{0,4}$	$r_{1,4}$

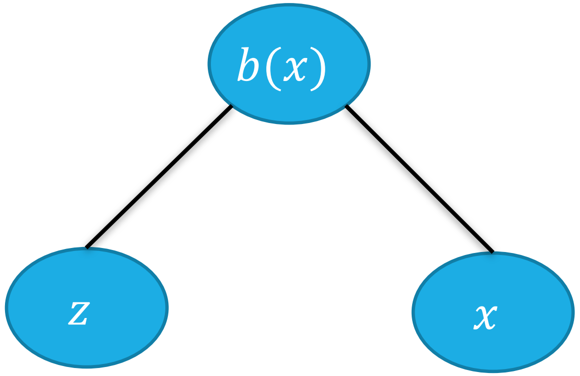
Causal effects are comparisons of $r_{1,i}$ *and* $r_{0,i}$*, such as* $r_{1,i} - r_{0,i}$ *or* $r_{1,i}/r_{0,i}$

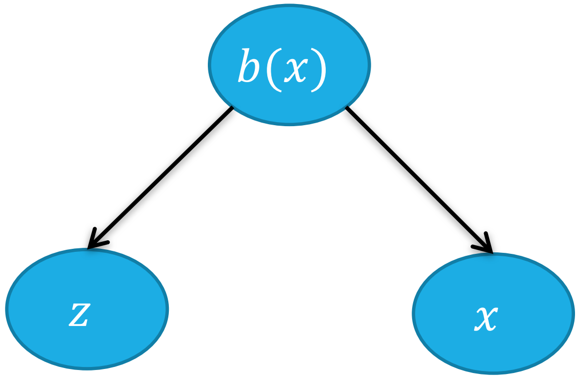
Estimating the causal effects of treatments is a missing data problem

The quantity to be estimated is the average treatment effect: $\mathbb{E}(r_1) - \mathbb{E}(r_0)$

BALANCING SCORES

A balancing score is a function $b(x)$ such that $p(z|x, b(x)) = p(z|b(x))$.





$x \perp z|b(x)$

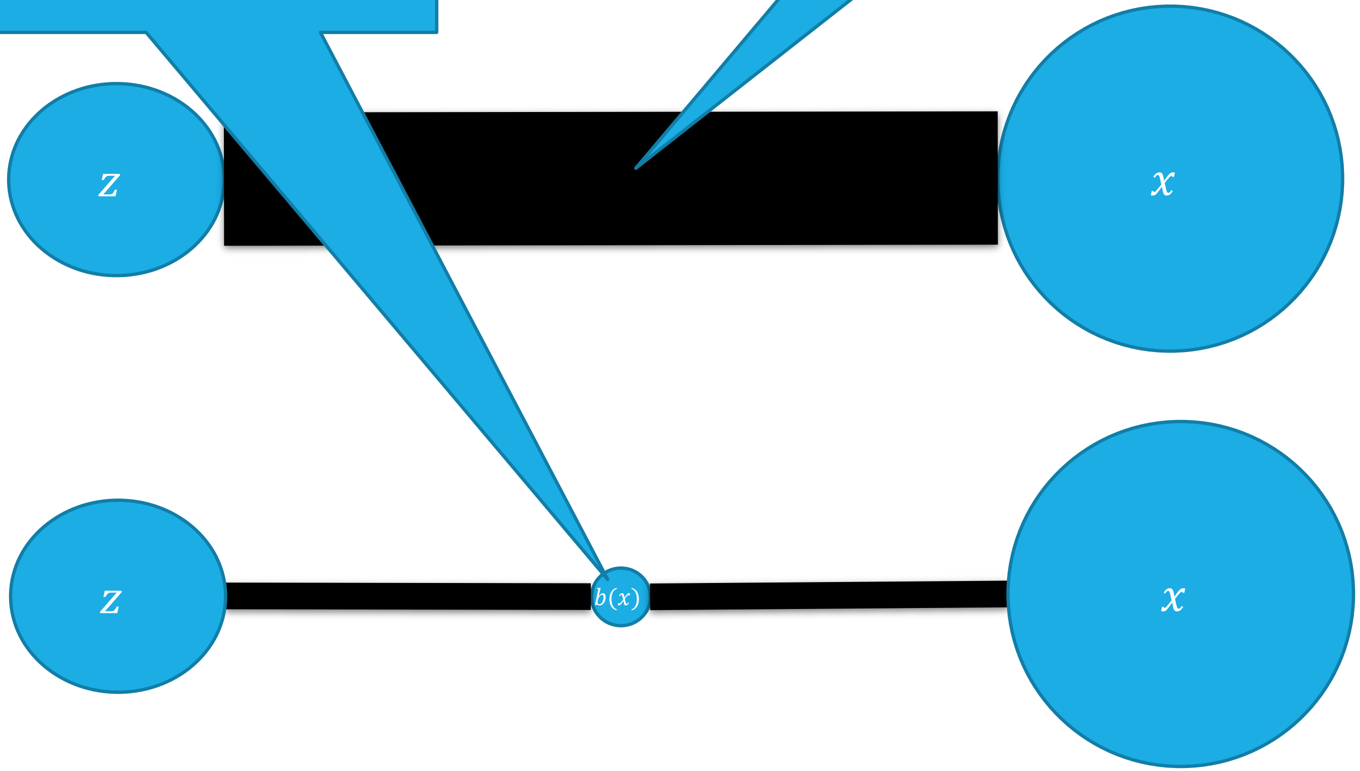
Intuitively $b(x)$ captures at least "all the information from x that is dependent on z"

It chops x into equivalence classes that are indistinguishable from the point of view of z

BALANCING SCORES

To model the relationship between x and z, we have to estimate a lot of parameters

A balancing score allows us to estimate far less



STRONGLY IGNORABLE TREATMENT ASSIGNMENT

Treatment assignment is **strongly ignorable** given a vector of covariates v, if

- Every unit in the population has a chance of receiving each treatment

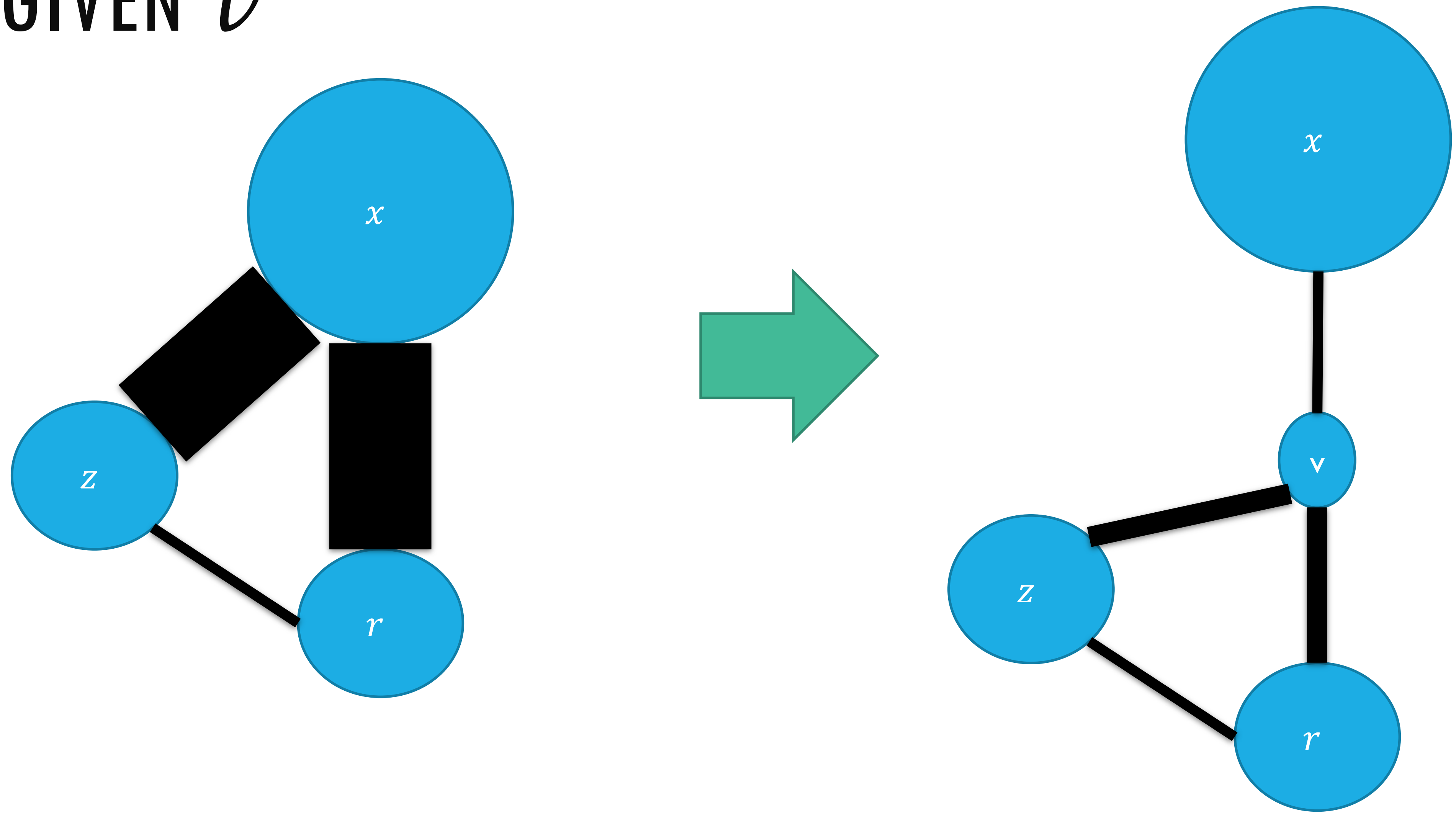
$$0 < p(z|v) < 1$$

- Treatment assignment and response are conditionally independent given v

$$(r_1, r_0) \perp z|v$$

Intuitively, there is no variable outside of v which has a correlation with the treatment and with the response, thus creating a dependence between them. That is, every relevant variable either for response or for treatment is included in v.

STRONGLY IGNORABLE TREATMENT ASSIGNMENT GIVEN v



ATE COMPUTATION BASED ON STRONGLY IGNORABLE TREATMENT ASSIGNMENT

If treatment assignment is **strongly ignorable** given v, then

$$\mathbb{E}(r_1|v, z = 1) - \mathbb{E}(r_0|v, z = 0) = \mathbb{E}(r_1|v) - \mathbb{E}(r_0|v)$$

and hence

$$\mathbb{E}_v\{\mathbb{E}(r_1|v, z = 1) - \mathbb{E}(r_0|v, z = 0)\} = \mathbb{E}_v\{\mathbb{E}(r_1|v) - \mathbb{E}(r_0|v)\} = \mathbb{E}(r_1) - \mathbb{E}(r_0) = ATE$$

Under strongly ignorable treatment assignment given v, units with the same values for v but different treatments can act as controls for each other.

Strongly ignorable treatment assignment given v, provides us a way to compute the ATE which is efficient if $v \ll x$

PROPENSITY SCORE

The propensity score $e(x)$ is the propensity of x towards exposure to treatment 1, or the probability of treating a unit with characteristics x

$$e(x) = p(z = 1|x)$$

The propensity score is known for randomized trials but unknown for nonrandomized trials. There, it may be estimated from observed data by means of, for example, logistic regression.

MAIN THEORETICAL RESULTS

Large sample theory ($e(x)$ is assumed to be known):

- The propensity score is the “thinnest” balancing score, and any score “thicker” than it is also a balancing score.
- If treatment assignment is strongly ignorable (given x):
  - It is strongly ignorable given any balancing score.
  - At any value of a balancing score, the difference between the treatment and control means is an unbiased estimate of the average treatment effect at that value of the balancing.

Small sample theory

- Using sample estimates of balancing scores can produce sample balance on x.

PRACTICAL METHODS. PAIR MATCHING ON BALANCING SCORES 1.0

When we have a small set of treated patients and a large reservoir of control patients

1. Select a balancing score $b(x)$
2. For each treated patient
    1. Compute the balancing score of the patient
    2. Sample a control unit with an equal (similar) balancing score.
    3. Compute the difference in response between the two units.
3. Compute the average all the differences computed in step 2.3

Use the average computed in step 2 as an estimate of the ATE.

PRACTICAL METHODS. PAIR MATCHING ON BALANCING SCORES 2.0

1. Select a balancing score $b(x)$ and determine its distribution on the population
2. Repeat
   - Sample a value α of the balancing score
   - Sample a treated unit and a control unit both with balancing score equal to α.
   - Compute the difference in response between the two units.
3. Compute the average all the differences computed in step 2.

The average computed in step 3 is unbiased for the ATE.

PRACTICAL METHODS. SUBCLASSIFICATION ON BALANCING SCORES

1. Sample a group of units such that
   1. $b(x)$ is constant in the group, and
   2. at least one unit in the group received each treatment.

The expected difference in treatment means equals the ATE at $b(x)$

The weighted average of such differences, when the weights equal the fraction of population at $b(x)$, is unbiased for the ATE

PRACTICAL METHODS. CAVEATS ON APPROXIMATING THE PROPENSITY SCORE

Usually the propensity score is unknown for an observational study.

We can estimate $e(x)$ by $\hat{e}(a) = prop(z = 1|x = a)$, the proportion in our sample.

Sample balance is only guaranteed for those values of $\hat{e}(a)$ satisfying $0 < \hat{e}(a) < 1$.

This raises **serious concerns** about these methods when the distribution over x does not have **finite support.**

In particular, subclassification on $\hat{e}(a)$ is guaranteed to work on large samples if x takes only finitely many values.

In practice, except when x takes on only a few values, $\hat{e}(a)$ will be either zero or one for most values of a. Consequently, in order to estimate propensity scores, some modelling will be required.

PRACTICAL METHODS. MODELING THE PROPENSITY SCORE

Different models impose specific parametric families on $p(x|z)$

Discriminant matching (Cochran & Rubin, 1973):

Assumes that

$$p(x|z = t) = N(\mu_t, \Omega)$$

Using logistic regression to estimate $p(z|x)$. This is assuming the logit model for $e(x)$.

Assumes that $p(x|z = t)$

$$p(x|z = t) = \frac{h(x) \exp w_t x}{f(w_t)}$$

SUMMARY

Balancing score = Captures the information from the covariates that is relevant to determine the treatment

Strongly ignorable treatment assignment = There is no relevant covariate for the determination of the treatment or of the response which is not included in our data sample.

Strongly ignorable treatment assignment given $v \ll x$, provides us an efficient way to compute the ATE

The propensity score $e(x)$ is the probability of treating a unit with characteristics x.

Strongly ignorable treatment assignment =>

Strongly ignorable treatment assignment given $e(x)$

$e(x)$ is the "thinnest" balancing score (look no further).

The theory provides three different practical methods for estimating ATE:

Matched samples, subclassification, covariance adjustment

QUESTIONS